

# The Interconnection of Willing and Believing in Kant's and Kantian Ethics

In this paper I look at the connection between willing and believing and, in particular, the role that this connection has in Kant's and Kantian ethics. I argue that the two main formulations of the categorical imperative are relativized to agents according to their beliefs, and I point out three different ways in which Kant or a present day Kantian might defend this. I conclude with some remarks about the contrast between Kant's legal theory and his ethical theory.

## 0 Introduction

Kant's ethics traditionally has been characterized as an ethics of willing, an ethics of maxims, an ethics of intention. These characterizations are meant to distinguish it from traditional forms of consequentialism, which are ethics of action, ethics of consequences, ethics of results.<sup>i</sup> What often gets overlooked is the deep connection between willing and believing and the shape that this gives to Kant's (and any plausibly Kantian) ethics.

One place where this connection can be seen is Kant's argument for the practical postulates.<sup>ii</sup> Although Kant's argument for the practical postulates changes in his various articulations of the doctrine of the highest good, the general scheme remains the same. Agents ought to promote the highest good as far as they are able. But agents cannot will something if they do not believe that the thing is really possible. That is, agents cannot will something unless they believe that it is not merely logically possible (i.e., free of internal contradiction) but also really possible (i.e., has a ground in actuality). The highest good is a real possibility only if God exists and agents are immortal. Therefore, agents, insofar as they are rational, ought to believe in God and in immortality.<sup>iii</sup> I take

this to show that Kant was at some level aware that what one wills can influence what one believes<sup>iv</sup> and even that the moral law prescribes certain sorts of belief.<sup>v</sup> The idea is not that one should accept some proposition X in the face of persistent and indisputable evidence to the contrary because of what one values. Rather, the idea is that in cases of epistemic uncertainty,<sup>vi</sup> what one believes can and sometimes should be influenced by what one values.

The connection between willing and believing goes in both directions. Kant gives an example of this in his discussion of the Scottish rebels who attempted to put Prince Stuart in command of the British throne in 1745-1746. He argues that some of the Scottish rebels believed that this uprising was their duty to the House of Stuart (RL, AA 06:333-334). The remainder of the passage indicates that Kant thinks that the Scottish rebels have mistaken honor for duty (honor is a mere simulacrum of duty). But the fact that the Scottish rebels believed that they had a duty to behave in a certain way led them to adopt certain maxims, perform certain actions, etc.<sup>vii</sup> More prosaic examples of the way in which belief can influence willing can be found by appeal to the ought implies can principle. If an agent believes that it is impossible to X, s/he probably will believe that s/he does not have a duty to X.

In this paper I am interested in the way beliefs can influence willing. I am interested in the way that an agent's beliefs can determine what maxims s/he adopts. In particular, I argue that unless one denies the problem of moral knowledge (to be explained more fully below), different agents can reach different results applying the same formulations of the Categorical Imperative (CI) to the same situations. In other words: *maxim + circumstances underdetermines the results of the tests proposed in the CI.*

The remainder of this paper is divided into three sections. First, I look at the Formula of Universal Law (FUL) and the Formula of Humanity (FH) in order to argue for my claim. Second, I give three different ways that Kant (and a present day Kantian) might deal with my claim (one of which is to deny the problem of moral knowledge). Finally, I conclude with some remarks about the difference between Kant's doctrine of right and his doctrine of virtue. I also point out that, depending on which way of dealing with my claim is accepted, the standard reading of Kant's ethics as an ethics of willing might be in need of some serious revision.

## **1 The problem of false beliefs**

I point out now that what I am calling the problem of false beliefs was not recognized explicitly by Kant as such. This gives rise to an exegetical dilemma, which I try to resolve by justifying this result in the next section in three different ways, all of which can be given a Kantian pedigree.

### **1.1 FUL: Simple conformity or through your will**

In his derivation of FUL in part I of the *Groundwork for a Metaphysics of Morals*, Kant argues as follows:

Since I have deprived the will of every impulse that could arise for it from obeying some law, nothing is left but the conformity of actions as such with universal law, which alone is to serve the will as its principle, that is, I ought never to act except in such a way that I also could will that my maxim should become a universal law. (GMS, AA 04:402)

There are two distinct principles articulated in this section of the text.<sup>viii</sup> The first is:

**L: Conform your actions to universal law.**

The second is:

**FUL: Act in such a way that you could will that your maxim should become a universal law.**

The main difference between L and FUL has to do with willing: FUL introduces the idea that one should be able to will one's maxim as a universal law. One's maxim must be able to become a universal law through one's will. This idea is not contained in L. L does not refer to what an agent can or cannot will; it simply tells an agent to conform his/her actions to universal law. L might be paraphrased as a claim about value. According to L, universal laws are unconditionally good in some sense. These two prescriptions come apart precisely because of the fact noted in the introduction: there is a connection between willing and believing. An example will make this clear.

Consider the maxim "to commit suicide from self-love when my prospects seem to afford more pain than pleasure." I can will this maxim to be a universal law just in case the maxim can become a universal law of nature through my will. Now if I believe that the purpose of self-love is to further life, then this maxim presumably cannot become a universal law of nature through my will, for (on reflection) I shall notice that as a universal law of nature this law would involve a contradiction: I am using self-love in a way inconsistent with its natural purposiveness. But if I believe that the purpose of self-love is to attain pleasure and to avoid pain, then this maxim can become a universal law of nature through my will. Thus, whether this maxim passes the test proposed in FUL (whether an agent can will that his/her maxim should become a universal law) will depend upon the agent's beliefs.

Someone might argue that there are two notions of willing in Kant. This idea can get off the ground by analogy to the two notions of wanting often at play in Plato's dialogues. Socrates sometimes argues with his interlocutors that they do not really want *that*. What they really want is *this*. For example, suppose that you believe that there is leftover Halloween candy in the lounge and you tell me, "I want to go into the lounge." I happen to know why it is that you want to go into the lounge. I also know that the leftover Halloween candy has been moved from the lounge to the copy room. So I say, "you do not want to go to the lounge *really*. *Really*, you want to go to the copy room."

There seem to be two senses of 'want' here, and probably there is a way in which both of us are correct. Just so, someone might argue that there are two senses of 'will' in play in Kant. There is what one rationally wills and what one wills (*simpliciter*). Continuing with the objection, the sense of 'will' at play in FUL is the former rather than the latter. Thus, FUL is independent of what an agent subjectively can will based on false beliefs. Thus (the objection concludes) my argument ((1) what agents can will depends on what they believe, (2) the results of FUL depend on what agents can will, therefore (3) the results of FUL depend on what agents can believe) rests on an equivocation.

But this objection rests on an incoherent interpretation of Kant. The whole point of FUL is that someone rationally can will maxim X if and only if maxim X can become a universal law through his/her will. The whole point of FUL is to make a bridge between normal willing and rational willing: FUL is the criterion of rational willing. So this objection would make FUL circular.

However, L is not relativized to the agent in this way. L contains no reference to what an agent can will, thus the fact that what an agent can will depends on what that

agent believes does not come into play. Regardless of what an agent believes about the purposiveness of self-love, there is a fact of the matter with regard to whether the suicide maxim can be a universal law. There is, of course, a fact of the matter about whether any particular agent can *will* the suicide maxim as a universal law. But the point is that this fact will depend on the agent's beliefs about the world. Thus, FUL and L can come apart. They can come apart because an agent might have false beliefs, and these false beliefs will inform what the agent can will.

Now what I am saying about FUL should not be misunderstood. The idea is not that an agent who convinces him/herself that s/he can will X as a universal law *eo ipso facto* can. Similarly, the idea is not that there is no fact of the matter with regard to whether an agent can will X as a universal law. There is a fact of the matter about whether an agent can will X as a universal law. The point is that this fact is not independent of the agent's beliefs about how the world works, where these might include things like beliefs about purposiveness, (folk) psychology, etc.

In sum: what one can will is based on what one believes about the world. An agent can will to do something that is logically impossible if s/he does not believe it to be so (consider Hobbes, who willed to square a circle). Similarly, what can become a universal law through one's will is based on what one believes about the world. If one's beliefs are false, then one might wind up adopting maxims that pass FUL but not L.<sup>ix</sup> Simply looking at an agent's maxim (and the circumstances in which the agent is situated) do not give enough information to get a result from FUL. Knowledge of the agent's beliefs must be taken into account in determining what can become a universal law through the agent's will.<sup>x</sup>

## 1.2 FH: Existing or expressing

In his derivation of FH in part II of the *Groundwork of a Metaphysics of Morals*, Kant articulates two principles.<sup>xi</sup> The first is:

**H: Rational nature exists as an end in itself.**<sup>xii</sup>

The second is:

**FH: So act that you use humanity, whether in your own person or the person of any other, always at the same time as an end, never merely as a means.**<sup>xiii</sup>

The difference between H and FH is similar to the gap between L and FUL.<sup>xiv</sup> H is an existence claim. H is a claim about things out there in the world and the value that they have. FH has to do with expressive reasons for acting.<sup>xv</sup> But one's expressive reasons for acting depend on one's beliefs about the world. If these beliefs are incorrect, then what one does might be in conformity with FH but not with H.<sup>xvi</sup>

An analogy will make this clear. Suppose I am talking with a friend. What I say will depend to some extent on what I believe about the meaning of words. When I misuse one particular word and my beliefs about the meaning of that word are readily identifiable, we can distinguish between what I (actually) said and what I intended to say. Similarly, FH latches on to expressive reasons for acting, thus it latches on to what I intend to express with my actions. H latches on to what I actually express with my actions. Although these two things (what I intend to express and what I actually express) usually cohere and inform one another in complex ways, there are cases in which they clearly come apart as a result of false beliefs.

Consider the action of engaging in sex apart from procreation. If I believe that the inclination to sex is an animal impulse and that its purpose is procreation, then upon reflection I shall come to believe that this action is immoral. The ground of proof for this would be that this kind of sex would constitute a surrendering of my personality. This kind of sex would be making use of my humanity as an instrument for the satisfaction of my animal lusts and desires.<sup>xvii</sup> But if I believe that the inclination to sex is inextricably bound up with my humanity and if I think its purpose is to maintain the hormonal balance required for physical and mental health, then in engaging in sex apart from procreation I would not be using my humanity as a mere instrument for the satisfaction of my animal lusts and desires. Rather, I would be trying to further my humanity as an end. My action would be expressive of the desire to harmonize with humanity as an end.

Because FH deals with what one is expressing in one's actions, whether one is acting in accordance with it will depend to some extent on one's beliefs. However, H is not relativized in this way. H contains no reference to acting or using, thus the fact that what an agent is expressing depends in part on the agent's beliefs does not come into play. Regardless of what the agent believes about the purposiveness of the inclination to sex, there is a fact of the matter: humanity exists as an end in itself. If an agent unintentionally infringes on his/her own rights or the rights of another, s/he has infringed on someone's rights regardless of whether his/her action is classified as innocent or morally reprehensible. The act itself constitutes the violation of H. But the violation of FH (if there was one) might be located elsewhere.

I bring up the discussion of rights because a distinction similar to the one I am drawing between H and FH can be seen in Kant's *Rechtslehre*. In the *Rechtslehre*, Kant



distinguishes between positive laws (laws that are, as a matter of fact, in place) and natural laws (laws that should be in place). An example to bring out this distinction is as follows. For many years, it was written into the US constitution that blacks count as only 0.6 of a person for the purposes of representation in congress.<sup>xviii</sup> This (among others) was a positive law that most people now recognize violated natural law. Someone who wanted to emphasize the legal analog of H or L would say that someone following the positive law is acting in a way that is legally culpable because the law is not a good one. Someone who wanted to emphasize the legal analog of FH or FUL would say that someone following the positive law is acting in a way that is not legally culpable, but would point out that this is an unfortunate result of a more fundamental error, *viz.*, a bad positive law.

The point is that what one expresses in one's actions is based on one's conception of the world. An agent can write a paper or a paragraph that winds up being total gibberish if s/he consistently but unintentionally misuses a word or phrase. The agent might have had a clear idea in mind; s/he might be able to articulate it very well once informed of the malapropism; one even might be able to figure out what s/he is trying to say from the context. But the meaning of the words s/he used and what s/he intended to express come apart. Similarly, whether one expresses respect for humanity or treats it as an end is based on what one believes about the world. If one's beliefs are false, then one might wind up performing actions that are in conformity with FH but not with H.

## **2 Bridging the gap: from intrinsic legality to intrinsic normativity and back**

In the previous section I argued for a distinction between L and H on the one hand

and FUL and FH on the other. This distinction is referred to by some commentators as the distinction between intrinsic legality and intrinsic normativity.<sup>xix</sup> But Kant does not recognize any distinction between L and H on the one hand and FUL and FH on the other. Thus, in this section I shall canvass three distinct ways of bridging the gaps between L and FUL and between H and FH. I shall give all three of them a Kantian pedigree. However, my goal is simply to show that there are these three positions. I do not want to argue that one is better than the other or that Kant recognized this problem clearly and distinctly and took an unambiguous position on it.

## **2.1 Denying the problem of moral knowledge: intuiting moral properties**

One way to bridge the gaps between L and FUL and H and FH is to deny the problem of moral knowledge. If one can intuit directly whether a maxim passes L or whether an action is in accordance with H, then no issue of false beliefs can arise. No issue of false beliefs can arise because one clearly and distinctly perceives whether a given maxim conforms to L or a given action conforms to H. Since one knows this directly, there can be no false beliefs that result in maxims that pass FUL but not L or actions that pass FH but not H. Thus, the gaps described in the previous section disappear.

There is evidence that Kant subscribes to a position like this. For example, in the *Critique of Practical Reason* Kant writes:

The most common understanding can distinguish without instruction what form in a maxim makes it fit for a giving of universal law and what does not. (KpV, AA 05:27)

And just a few pages later, Kant makes the following, similar claim:

What is required in accordance with the principle of autonomy of choice is easily and without hesitation seen by the commonest intelligence... That is to say, what duty is, is plain of itself to everyone... (KpV, AA 05:38)

These two passages indicate that Kant thinks that moral knowledge is unreflective (known “without hesitation”), universal (known by the “commonest intelligence,” the “most common intelligence” and “everyone”) and not acquired with difficulty (known “easily,” for it is “plain of itself”).

Hints of this view are also prominent in the *Groundwork for a Metaphysics of Morals*. In part I of the *Groundwork for a Metaphysics of Morals*, Kant argues that if an action is from duty, it *eo ipso facto* is in conformity with duty (GMS, AA 04:390).<sup>xx</sup> The idea behind this conditional seems to be the following: (1) if an agent is acting from duty, s/he must believe that the action is a duty and (2) if some action is a duty, then it is in conformity with duty. But this line of reasoning gets us to the desired conclusion (that if an action is from duty, it *eo ipso facto* is in conformity with duty) only if Kant also is assuming that agents cannot be incorrect in believing that an action is a duty. That is, this line of reasoning goes through only if one denies the problem of moral knowledge, for if an agent can be incorrect in believing that an action is a duty, then it is not the case that if an agent believes that some action is a duty, then that action is a duty.

In other words, unless one denies the problem of moral knowledge, it could be the case that some agent believes that some action is a duty even though that action is contrary to duty. But such an agent could act from duty even though the agent’s action might not be in conformity with duty. Thus, unless one denies the problem of moral knowledge, the claim at GMS, AA 04:390 breaks down.

The point is not that it is implausible to argue that if an action is from duty, it *eo ipso facto* is in conformity with duty. The point is that Kant’s claim at GMS, AA 04:390

reveals something. In particular, it reveals that Kant seems to be committed to a denial of the problem of moral knowledge. Just as the passages from the *Critique of Practical Reason* indicate, Kant, at times, seems to be committed to the idea that agents have an infallible and immediate grasp of moral truths. Whatever the philosophical repercussions of such a view might be, it hopefully is clear that there is evidence that Kant was tempted by it. And hopefully it also is clear that it would bridge the gap between L and H on the one hand and FUL and FH on the other.

## **2.2 Denying the relevance of legality: conscience as the inner judge**

Another way to bridge the gaps between L and FUL and H and FH is to argue that, in ethics, one is one's own judge. If one has acted in accordance with one's own moral principles, then, for the purposes of ethics, one has done all that one ought. The issue of false beliefs becomes irrelevant because insofar as one makes L or H the principle of one's will, one is guiltless, regardless of whether one's (subordinate) maxims really could be universal laws or really are in accordance with the existence of rational nature as an end in itself. If one tries to conform one's maxims to universal laws or if one tries to act in accordance with the existence of rational nature as an end in itself, that is enough for the purposes of ethics. (It might not be enough for right, but that is a different story.)

There is evidence that Kant subscribes to a position like this. Kant makes the following claim in the second half of the *Metaphysics of Morals*:

But if someone is aware that he has acted in accordance with his conscience, then as far as guilt or innocence is concerned nothing more can be required of him. (TL, AA 06:401)<sup>xxi</sup>

A thorough discussion of Kant's notion of conscience would be too difficult to undertake here. But in order to understand the relevance of Kant's claim at TL, AA 06:401, one must know what roles Kant takes conscience to fulfill. Kant thinks that conscience acts as one's inner judge in determining whether one has behaved morally.<sup>xxii</sup> As such, conscience has two roles. An agent's conscience judges whether s/he has fulfilled his/her first-order duties. In addition, an agent's conscience judges whether s/he has exercised due care in determining whether a certain action is permissible (i.e., whether the agent has been negligent).<sup>xxiii</sup> In other words, conscience judges (1) whether an agent has done what s/he ought and (2) whether an agent has exercised due care in figuring out what s/he ought to do.

Now the whole point of discussing an agent who adopts a bad principle as a result of (not despite) a failure in judgment or a false belief is that such an agent seems to be acting in accordance with conscience. There is no inner judge that could condemn such an agent at the time of acting; indeed, the inner judge approves of the action. The agent (*ex hypothesi*) meets the internal conditions of good willing. The quotation from TL, AA 06:401 suggests that Kant held the view that such agents are inculpable.

Someone might object that the quotation at TL, AA 06:401 does not suggest this at all. The idea behind such an objection might be as follows. Strictly speaking, the idea that acting in accordance with conscience is sufficient for moral guiltlessness is consistent with (and perhaps bolstered by) holding that error about whether, for example, a maxim is universalizable is impossible without systematic self-deception. The objector might go on to point out that, indeed, in the very passage in which Kant claims that acting in accordance with conscience is a sufficient condition for moral guiltlessness, Kant claims

that conscience cannot err.

However, this objection does not withstand a close scrutiny of the text. The argument for the claim that conscience cannot err is as follows:

...while I can indeed be mistaken at times in my objective judgment as to whether something is a duty or not, I cannot be mistaken in my subjective judgment as to whether I have submitted it to practical reason (here in its role as judge) for such a judgment; for if I could be mistaken in that, I would have made no practical judgment at all, and in that case there would be neither truth nor error. (TL, AA 06:401)<sup>xxiv</sup>

There are two things of note about this argument for the claim that conscience cannot err.

First, the argument for the claim that conscience cannot err does not establish that conscience cannot err either in its role in assessing whether an agent has fulfilled first order duties or in its role in assessing whether an agent has exerted due diligence in determining whether a given action is permissible. Rather, the argument is intended to establish that “I cannot be mistaken in my subjective judgment as to whether I have submitted it [i.e., the action] to practical reason (here in its role as judge) for such a judgment.” This is analogous to the role of the bell on a microwave. I shall call this the check function. Conscience, in this capacity, tells an agent whether s/he has submitted an action that s/he is considering to the principles that s/he holds. The check function gives a yes or no answer. The check function does not tell an agent whether the action is in accordance (or not in accordance) with the agent’s moral principles. Thus, it does not tell an agent whether s/he has performed his/her first order duties. The check function tells an agent whether s/he has asked the question, “Have I performed my first order duties?” The check function does not tell an agent what the answer to this question is. Thus, the check function is not equivalent to the function conscience fulfills in telling an agent whether s/he has fulfilled his/her first order duties.<sup>xxv</sup> Similarly, the check function is not

equivalent to figuring out whether an agent has exerted due diligence in determining whether a given action is permissible. It is only once the check function gives a positive answer that one is in a position to ask whether one has engaged in due diligence.

Second, Kant admits explicitly in this passage that “I can indeed be mistaken at times in my objective judgment as to whether something is a duty or not.”<sup>xxvi</sup> Given that he admits this in the same passage in which he claims that acting in accordance with conscience is a sufficient condition for moral guiltlessness, it seems to me to be reasonable to conclude that some parts of Kant’s corpus suggest (1) that agents can make mistakes about whether a maxim conforms to universal law or whether an action conforms to the existence of rational nature as an end in itself<sup>xxvii</sup> and (2) agents who make such mistakes are guiltless in their adoption of such maxims or performance of such actions.<sup>xxviii</sup>

### **2.3 A middle ground: distinguishing between natural and acquired obligations**

Arguing that agents never can be incorrect about their obligations because they have an immediate intuition of the moral properties of their maxims and actions is an extreme position. Arguing that agents can be incorrect about their obligations but that this is inconsequential (for moral purposes) so long as they act in accordance with their consciences (in accordance with what they take their obligations to be) is also an extreme position. However, there is an intermediate position suggested by some passages in Kant’s lectures on ethics.

At V-Mo/Collins, AA 27:355 Kant argues:

In regard to his natural obligations, nobody can be in error; for the natural moral laws cannot

be unknown to anyone, in that they lie in reason for all; hence nobody is guiltless there in such error. (V-Mo/Collins, AA 27:355)

In this passage Kant suggests that there are some cases (“natural obligations”) in which there is no problem of moral knowledge. In these cases, no gap can open up between L and FUL or H and FH. But in a similar passage Kant claims that distinguishing between natural obligations and acquired obligations (non-natural obligations) is notoriously difficult (V-Mo/Herder, AA 27:42-43).<sup>xxix</sup> There are two implications of Kant’s discussion: (1) An agent can be guiltless in an error about acquired obligations and (2) it is very difficult to distinguish between natural obligations and acquired obligations.<sup>xxx</sup> (1) indicates that there are some cases (acquired obligations) in which there is a gap between L and FUL and H and FH. (2) suggests that even in cases of natural obligations, this gap might *seem* to open up. The gap might seem to open up because it is very difficult to distinguish between natural obligations and acquired obligations.

The distinction between natural and acquired obligations must be taken as a starting point. The distinction might be vague at the borders, but that does not prevent one from giving examples of what Kant probably would have thought fell under these headings. An example of an infringement of an acquired obligation would be a person who chews with his/her mouth open in a restaurant. An example of an infringement of a natural obligation would be the actions of Himmler and his S.S. officers. Indeed, Himmler himself seems dimly aware of the fact that he and his officers were infringing on a natural obligation. In a speech made to S. S. officers he is reported to have said the following:

...I also want to talk to you quite frankly on a very grave matter... I mean... the extermination of the Jewish race... Most of you must know what it means when 100 corpses are lying side by side, or 500, or 1,000. To have stuck it out and at the same time — apart from exceptions caused by human weakness — to have remained decent fellows, that is what has made us hard.<sup>xxxi</sup>



Himmler boasted that the S. S. had succeeding in wiping out the Jews

...without our leaders and their men suffering any damage in their minds and souls. The danger was considerable, for there was only a narrow path between the Scylla of their becoming heartless ruffians unable any longer to treasure life, and the Charybdis of their becoming soft and suffering nervous breakdowns.<sup>xxxii</sup>

This, I submit, is the talk of somebody who knows that what he is doing is wrong. He admits the difficulty in remaining “decent fellows” in the face of 100, 500 or 1,000 corpses lying side by side. The S. S. officers had to steer a path between becoming “heartless ruffians” on the one side and suffering “nervous breakdowns” on the other. One simple explanation for this is that they were going against the grain and they knew it.

Of course, Himmler would not be happy with this explanation. Himmler would say that he is forced (by moral compulsion) to carry out a disagreeable duty. Natural sympathy might make it pain a judge that he must condemn a guilty man; a physician might abhor the sight of blood when he must cut a patient in surgery to save his life. The fact that they feel this ambivalence, which could be characterized as a conflict between “hard” duty and a bad tendency to be “soft,” does not show that they are doing wrong in performing their duties as a judge or surgeon. Himmler would try to rationalize his conflicts by comparing them to cases like this.

The point of the discussion, however, is not to take a stand on whether there are natural obligations.<sup>xxxiii</sup> The point is that if there are natural obligations, then Himmler, surely, is someone who infringed upon them; the thin layer of self-deception present in his speeches and memoirs seems to provide evidence for this. But I do not need (and am not going to try) to give some sort of apodictic proof that Himmler was aware that what he was doing was wrong. The point is that *if* there are such things as natural obligations, then surely Himmler is someone who infringed upon them.

The distinction between natural and acquired obligations would bridge the gaps between H and FH and L and FUL in different ways in different cases. In the case of natural obligations, the gaps would not open up because (*ex hypothesi*) agents cannot be incorrect inculpably about such obligations. Agents could not be incorrect inculpably about whether maxims associated with natural obligations are universalizable or whether actions associated with natural obligations are in conformity with the existence of humanity as an end in itself. However, in cases of acquired obligations, agents can be mistaken about these things. And the implication of the passages from Kant's lectures is that in such cases, agents might be guiltless. Thus, in cases of acquired obligation in which agents are mistaken about the properties of their maxims or their actions and thus are acting in conformity with FUL or FH but not with L or H, the agents are acting inculpably.<sup>xxxiv</sup>

This position incorporates aspects of both of the positions articulated in the previous two subsections. Perhaps somebody more ambitious than I could massage the texts in such a way as to show that Kant has one clear and coherent position. (If so, Kant's position most likely would take the distinction between natural and acquired obligations as a starting point as suggested in this subsection. The quotations from the first subsection would be taken to be overstatements of Kant's position in one direction; agents do not always know immediately whether their maxims are universalizable. The quotations from the second subsection would be taken to be overstatements of Kant's position in the other direction; agents can make nonculpable mistakes about their obligations only in the case of acquired obligations.) But my goal has been merely to articulate the different options and to show how they interrelate.

### 3 Conclusion

In this paper I argued that there is a gap between L and H on the one hand and FUL and FH on the other. This gap arises because what one wills and what one expresses are dependent upon one's beliefs, and one's beliefs might be false. I argued that this gap can be closed in one of three ways. First, someone could deny the problem of moral knowledge: someone could argue that agents have a direct, clear, distinct and intuitive knowledge of the moral properties of maxims and actions. Second, someone could argue that if one is acting in accordance with one's moral principles as one accepts them, then one is guiltless as far as morality is concerned. Thus, if one makes L or H the principle of one's will, then one has done all that one ought, false beliefs and bad actions notwithstanding. It might be an intrinsic feature of maxims that they have lawlike form or that they are in conformity with humanity as an end in itself. But to determine whether an agent is acting in accordance with the Categorical Imperative is a difficult question, and figuring out the intrinsic properties of an agent's maxim or action does not exhaust it. Third, someone might make a distinction between natural and acquired obligations and argue that, in the case of the former, agents cannot be mistaken inculpably whereas in the case of the latter, agents can. If one accepts any but the first option, then the standard reading of Kant's as an ethics of willing, an ethics of maxims, an ethics of intention is in need of some serious revision.

Before I conclude, I shall say something about the distinction between right and good on Kant's theory. According to Kant's theory of right, the privileged perspective is

not that of the agent but that of the judge who is taken to be a representative of the general will. A judge in a court of law might decide whether an infringement of right has occurred by appealing to an agent's intents, beliefs and maxims more broadly.

Thus, regardless of which position one accepts (i.e., regardless of whether one denies the problem of moral knowledge or not), according to a Kantian theory of right agents can be punished rightfully and can be judged legally culpable for doing things regardless of their intentions. If one does deny the problem of moral knowledge, then a certain view of the interconnection between right and good can arise (*viz.*, a monist view according to which legal culpability implies moral culpability). If one does not deny the problem of moral knowledge, then another view can arise (*viz.*, a dualist view according to which legal culpability does not imply moral culpability). But the questions that this raises are far beyond the scope of this paper.

---

<sup>i</sup>See, e.g., Barbara Herman, *The Practice of Moral Judgment* (Cambridge MA: Harvard University Press, 1993), p. 94p1: "Kantian ethics is moral theory based on intention or volition. The objects of moral assessment are not events or states of affairs, but willings (or willed actions). Things that happen are not themselves morally good or bad, right or wrong: only willings are. If you take this seriously (and I take it as defining of Kantian ethics)...". I discuss this at greater length in my Samuel Kahn "A reply to Bencivenga, "Consequences in Kantian Ethics."" *American Dialectic* III.1 (January 2013): 285-288.

<sup>ii</sup>This argument is articulated at various places throughout Kant's corpus, including the *Critique of Pure Reason*, the *Critique of Practical Reason* and the *Critique of the Power of Judgment*.

<sup>iii</sup>Cf., e.g., Paul Guyer, *Kant on Freedom, Law, and Happiness* (Cambridge UK: Cambridge University Press, 2000), chapter 10 or Allen Wood, *Kant's Moral Religion* (Ithaca NY: Cornell University Press, 1970).

<sup>iv</sup>Here I am taking *vernunftige Glaube* (often translated as 'faith') to be a kind of belief.

<sup>v</sup>Another example of this might be the belief in different temperaments. One argument for this might be as follows.

Agents have a duty to believe that most people are morally decent. But when one looks around in the world, one can see that different people have achieved different results with regard to their moral proclivities. If one accepts that most people are morally decent and if one accepts that different people have achieved different results with regard to their moral proclivities, then it seems reasonable to infer that different people have different temperaments. In other words, our moral identities are informed not merely by nurture (self-nurture as well as the nurture of our parents, peers and others); our moral identities are informed also by our natures. Whether we are responsible for all that results from our natures seems to be an independent question.

<sup>vi</sup>In the case of the postulates, the epistemic uncertainty is a result of our faculties; Kant insists that human agents cannot have cognition about the content of the practical postulates. However, Kant's remarks at the end of the *Critique of Pure Reason* about degrees of belief as well as his remarks in the *Religion within the Boundaries of mere Reason* about probabilism suggest that he thinks that practical considerations can (and

---

perhaps should) influence belief even in cases about which humans, at least theoretically, could have cognition.

<sup>vii</sup>Kant's discussion of mothers who commit infanticide when they have babies out of wedlock and officers who engage in duels when they have been insulted could be similar if such agents believe that they have a duty to engage in these activities (RL, AA 06:336-337).

<sup>viii</sup>The gap that I am about to point out can be found also in Kant's derivation of FUL in part II of the *Groundwork for a Metaphysics of Morals* and in the derivation of FUL in the *Critique of Practical Reason*. Cf. Bruce Aune, *Kant's Theory of Morals* (Princeton NJ: Princeton University Press, 1980), esp. p. 29p1-2 and Allen Wood, *Kant's Ethical Thought* (Cambridge UK: Cambridge University Press, 1999), esp. pp. 78p2-79p2. Although I believe that the discussion in this paper is relevant to discussions of the derivation of FUL, I cannot pursue that line of thought here.

<sup>ix</sup>I do not mean to be ruling out the possibility of adopting maxims that pass L but not FUL. But discussion of this possibility is more complicated.

<sup>x</sup>Someone might object that since belief-forming is an act, it is assessable by the universalizability tests, hence the gap between L and FUL is narrower than might appear. But this misses the point. The point is that after those beliefs have been formed, the moral landscape changes, for the beliefs will determine partly what an agent can will as a universal law (regardless of whether the beliefs are culpably incorrect).

<sup>xi</sup>As in the previous subsection, I believe that the distinction I am drawing between H and FH is relevant to discussion of the derivation of FH. But I cannot pursue this line of thought here.

<sup>xii</sup>GMS, AA 04:428-429.

<sup>xiii</sup>GMS, AA 04:429.

<sup>xiv</sup>As pointed out in the footnotes above, the gap between L and FUL is widely accepted. However, the gap between

H and FH is not. In fact, the only commentators whom I know of who are sensitive to the gap between H and FH are Wood and Herman (see Wood, *Kant's Ethical Thought*, chapter 5 and Herman, p. 87n12, respectively). Wood seems to make this gap one of the cornerstones of his criticism of constructivist readings of Kant. But to discuss these issues here would be too difficult.

<sup>xv</sup>For a discussion of the so-called regress argument for the Formula of Humanity and, in particular, the idea that agents necessarily represent their ends as objectively good, see Samuel Kahn, "The guise of the objectively good," *The Journal of Value Inquiry*, 2013, 10.1007/s10790-013-9372-2.

<sup>xvi</sup>As in the previous section, I do not mean to deny that agents with false beliefs can act in conformity with H but not with FH.

<sup>xvii</sup>For a discussion of the varied ways in which Kant uses the term 'humanity' see Samuel Kahn, "Reconsidering 6:26n and the meaning of 'Humanity,'" in Stefano Bacin, Alfredo Ferrarin, Claudio La Rocca, and Margit Ruffing (ed.'s), *Proceedings of the XIth International Kant Congress* (Berlin Germany: Walter de Gruyter).

<sup>xviii</sup>Technically, the so-called Three-Fifths Compromise was about "all other persons," other from those who were free, from those bound in service and from Indians who did not have to pay taxes.

<sup>xix</sup>See esp. Christine Korsgaard, *Creating the Kingdom of Ends* (Cambridge UK: Cambridge University Press, 1996), pp. 61p1-67p2.

<sup>xx</sup>Herman has dubbed this the non-accidental rightness condition (Herman, chapter 1).

<sup>xxi</sup>Cf. also, e.g., V-Mo/Collins, AA 27:335 and 355.

<sup>xxii</sup>See, e.g., KpV, AA 05:98; TL, AA 06:438; V-Mo/Collins, AA 27:295, 351 and 353-354; and V-Mo/Vigilantius, AA 27:572-573.

<sup>xxiii</sup>See, e.g., RGV, AA 06:186: "Conscience does not pass judgment upon actions as cases that stand under the law, for this is what reason does so far as it is subjectively practical... rather, here reason judges itself, whether it has actually undertaken, with all diligence, that examination of actions (whether they are right or wrong), and it calls upon the human being himself to witness for or against himself whether this has taken place or not." Cf., e.g., V-Mo/Vigilantius, AA 27:614-615.

<sup>xxiv</sup>The textual difficulties here are overwhelming. At one point Kant argues that conscience cannot err on the grounds that conscience is an instinct, thus must be distinguished from the faculty of speculative judgment because it is not free (V-Mo/Collins, AA 27:297). At another point it seems that Kant takes the claim that conscience cannot err as a starting point in order to adduce its function (V-Mo/Vigilantius, AA 27:615). However, at places where Kant attributes a wider role to conscience than he does at TL, AA 06:401, he admits explicitly that conscience can err (see, e.g., V-Mo/Herder, AA 27:42 and V-Mo/Collins,

---

AA 354-355). Indeed, some kinds of moral errors that Kant describes do seem to be errors of conscience. For example, at GMS, AA 04:405 Kant claims that sometimes agents judge themselves innocent when they are not, deceiving themselves by applying too lenient standards. And at TL, AA 06:356-357 Kant claims that sometimes agents judge themselves too harshly (he calls this a “morbid conscience”). It is not clear why these errors could not be called errors of conscience. In light of these difficulties, I hope I shall be excused for confining myself to the passage at TL, AA 06:401. For a more detailed discussion, see my Samuel Kahn, “Kant’s theory of conscience,” in Pablo Muchnik and Oliver Thorndike (ed.’s), *Rethinking Kant: Volume IV*, (Cambridge UK: Cambridge Scholars Publishing).

<sup>xxv</sup>Here I am disagreeing with Hill, who argues that Kant’s claim that conscience cannot err is false on the grounds that agents do make mistakes in matching up their acts (as they perceive them) with their moral principles (Thomas Hill, *Human Welfare and Moral Worth: Kantian Perspectives* (Oxford UK: Oxford University Press, 2002), p. 348p3). If my reading of Kant’s claim is correct, then Hill’s argument against the passage at TL, AA 06:401 does not connect. I discuss this in more detail in Samuel Kahn, “Kant’s theory of conscience,” in Pablo Muchnik and Oliver Thorndike (ed.’s), *Rethinking Kant: Volume IV*, (Cambridge UK: Cambridge Scholars Publishing).

<sup>xxvi</sup>Kant makes a similar claim at Anth, AA 07:292: “though it is true that these principles occasionally might be mistaken and imperfect, still the formal element of his volition in general — to act according to firm principles... — has something precious and admirable in it...”

<sup>xxvii</sup>Kant himself made some very serious errors about who counts as a rational being and about what kinds of maxims conform to universal laws.

<sup>xxviii</sup>Recall Raskolnikov’s confession: “...you must not suppose that I did not know, for instance, that if I began to question myself whether I had the right to gain power — I certainly had not the right — or that if I asked myself whether a human being is a louse it proved that it was not so for me, though it might be for a man who would go straight to his goal without asking questions.”

<sup>xxix</sup>The difficulty in distinguishing between natural obligations and acquired obligations seems to be analogous to the difficulty associated with distinguishing between the *a priori* and the empirical in the domain of theoretical knowledge (KrV, AA B1-2, A23/B36).

<sup>xxx</sup>I should point out that Kant does not use the term ‘acquired obligations’. In the passage at V-Mo/Collins, AA 27:355, he contrasts “natural obligations,” which arise from natural laws, with obligations that arise from positive laws. He also distinguishes between natural conscience and artificial conscience. At V-Mo/Herder, AA 27:42-43, he distinguishes between natural conscience and acquired conscience. (Obviously) I am using “natural obligations” to refer to the obligations that are in accordance with one’s natural conscience and “acquired obligations” to refer to the obligations in accordance with one’s “acquired conscience.” The distinction that is operative in the text above, which should be clear, is the distinction between obligations in regard to which one inculpably can be in error and obligations in regard to which one cannot be in error inculpably.

<sup>xxxi</sup>William Shirer, *The Rise and Fall of the Third Reich* (New York NY: Simon & Schuster, 1960), p. 966.

<sup>xxxii</sup>Heinrich Fraenkel and Roger Manwell, *Heinrich Himmler* (London UK: New English Library, 1965), p. 132.

<sup>xxxiii</sup>Recall O’Brien’s chilling words: “you are imagining that there is something called human nature which will be outraged by what we do and will turn against us. But we create human nature. Men are infinitely malleable.”

<sup>xxxiv</sup>Strictly speaking, I should say merely that such agents *might* be acting inculpably. But hopefully the idea is clear.